Recent Developments in the Use of Clinical Trials to Support Individualizing Therapies:

A Regulatory Perspective

Robert T. O'Neill Ph.D.

Director, Office of Biostatistics

CDER ,FDA

For presentation at the 4th Seattle Symposium in Biostatistics: Clinical Trials, November 20-23, 2010

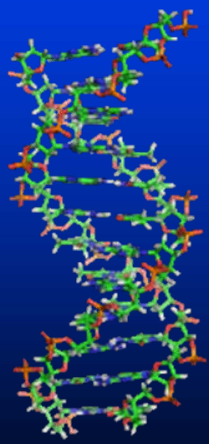
Outline

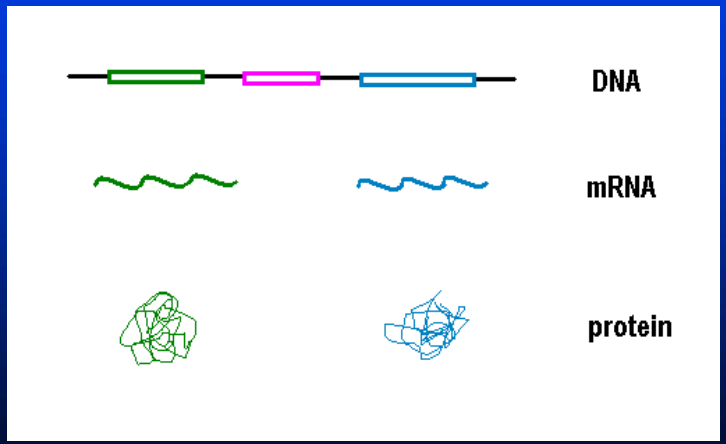
- What is meant by individualized therapy
- Some history of the statistical interest in this issue
- Some current clinical trial experience
- Relationship to ICH E5 – Acceptance of foreign clinical data
- Challenges for the randomized trial and product development
- Where are we going

Personalized (Individual) Medicine

What does it mean ?

Biomarkers and Classifiers





What does individualized therapy mean

- If you cannot metabolize a drug, meaning the drug will not have its intended pharmacological effect(s) then you cannot benefit from the drug and may just share its risk or side effects.
- If you are a slow, intermediate or fast metabolizer of a drug , you may need a different dose of a drug to get a comparable effect
- If the target of the drug is resistant or non-responsive to the therapy, then the intended therapeutic effect is neutralized or minimized
- If you have the marker(s) you should get a better response to the treatment in contrast to a patient without the marker

There is a rich statistical history of indentifying prognostic factors

Int. J. Cancer: **13,** 16-36 (1974)

STATISTICAL METHODS FOR THE IDENTIFICATION AND USE OF PROGNOSTIC FACTORS [1]

by

P. ARMITAGE [2] and Edmund A. GEHAN [3]

[2] *Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, England;*
[3] *Department of Biomathematics, The University of Texas M. D. Anderson Hospital and Tumor Institute at Houston, Houston, Texas 77025, USA*

Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually

Vol. 81, No. 24, December 20, 1989

Journal of the National Cancer Institute

Mitchell H. Ga P. Byar, Donald K. Corle, Sylvan B. Green, Catherine Schairer, John J. Mulvihill

The Choice of Treatment for Cancer Patients Based on Covariate Information :

Application to Prostate Cancer

Bull Cancer (Paris), 1980, 67, 4, 477-490.

D. P. BYAR and S. B. GREEN

Clinical and Diagnostic Trials Section, Biometry Branch, National Cancer Institute, Bethesda.

Eur J Cancer Clin Oncol, Vol. 17, No. 12, pp. 1307–1313, 1981.
Printed in Great Britain.

0277–5379/81/121307–07$02.00/0

Prognostic Significance of CEA in Breast Cancer: a Statistical Study*

MAARTJE DE JONG-BAKKER,† AUGUSTINUS A. M. HART,† JEAN-PAUL PERSIJN‡§ and FRANS J. CLETON†

Departments of Medicine† and Clinical Chemistry,‡ Antoni van Leeuwenhoek-Huis/Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

STATISTICS AND MEDICINE

The Limitations of Risk Factors as Prognostic Tools

James H. Ware, Ph.D.

Related article, p. 2631

American Journal of Epidemiology

Vol. 159, No. 9
Printed in U.S.A.
DOI: 10.1093/aje/kwh101

Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker

Margaret Sullivan Pepe[1,2], Holly Janes[2], Gary Longton[1], Wendy Leisenring[1,2,3], and Polly Newcomb[1]

[1] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.
[2] Department of Biostatistics, University of Washington, Seattle, WA.
[3] Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA.

Received for publication June 24, 2003; accepted for publication October 28, 2003.

The NEW ENGLAND JOURNAL *of* MEDICINE
N ENGL J MED 355;25 WWW.NEJM.ORG DECEMBER 21, 2006

ORIGINAL ARTICLE

Multiple Biomarkers for the Prediction of First Major Cardiovascular Events and Death

Thomas J. Wang, M.D., Philimon Gona, Ph.D., Martin G. Larson, Sc.D., Geoffrey H. Tofler, M.D., Daniel Levy, M.D., Christopher Newton-Cheh, M.D., M.P.H., Paul F. Jacques, D.Sc., Nader Rifai, Ph.D., Jacob Selhub, Ph.D., Sander J. Robins, M.D., Emelia J. Benjamin, M.D., Sc.M., Ralph B. D'Agostino, Ph.D., and Ramachandran S. Vasan, M.D.

What is new with prediction ?

- Predicting the treatment effect (compared to what) not the clinical outcome itself (single cohort idea)
- Enrichment designs
- Adaptive designs
- Type 1 error control for multiple subgroup hypotheses
- Biomarkers as classifiers and their validation (qualification)


VOLUME 27 · NUMBER 24 · AUGUST 20 2009

JOURNAL OF CLINICAL ONCOLOGY

STATISTICS IN ONCOLOGY

Clinical Trial Designs for Predictive Biomarker Validation: Theoretical Considerations and Practical Challenges

Sumithra J. Mandrekar and Daniel J. Sargent

What appears different about targeted therapy designs

- Not framed as covariates as prognostic factors
- Not framed as a subgroup problem with the need for statistical interaction tests (known to be of low power against most alternatives)
- Differential treatment response as a function of predictive factors
- Study design implication: Multiple hypotheses
  - Allocate type 1 error to several hypotheses of interest , including the all comers and a targeted subset

How does ICH E5 (Acceptance of foreign clinical data) apply

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Food and Drug Administration

[Docket No. 97D–0299]

International Conference on Harmonisation; Guidance on Ethnic Factors in the Acceptability of Foreign Clinical Data; Availability

Federal Register / Vol. 63, No. 111 / Wednesday, June 10, 1998 / Notices

Key Features of E5

- Ethnic factors classified as extrinsic and intrinsic
  - Does the effect of a drug differ by these factors
- Introduced concept of multi-regional clinical trial to support evidence in different regulatory regions – the bridging concept
- Provided a cap on how much additional data could be asked for by a regulator in a region
  - Allowed another clinical study to be requested if needed

Classification of intrinsic and extrinsic factors

Appendix A: Classification of intrinsic and extrinsic ethnic factors

INTRINSIC		EXTRINSIC
Genetic	Physiological and pathological conditions	Environmental
Gender	Age (children-elderly)	Climate
Height		Sunlight
Bodyweight		Pollution
	Liver	**Culture**
	Kidney	Socioeconomic factors
	Cardiovascular functions	Educational status
ADME		Language
Receptor sensitivity		
Race		Medical practice
		Disease definition/Diagnostic
Genetic polymorphism of the drug metabolism		Therapeutic approach
		Drug compliance
		Smoking
		Alcohol
		Food habits
Genetic diseases	Diseases	Stress
		Regulatory practice/GCP
		Methodology/Endpoints

The intrinsic factor focus
Current interest

- Pharmacogenomics
  - Relating genomic profiles to differential benefit or risk or dosing paradigms
  - Evaluating the differential prevalence of such genomic profiles for different ethnic and racial groups to determine dependency of dose, benefit / risk on profiles

The Q & A addendum

Introduced the concept of multi-regional RCT for bridging

Guidance for Industry

E5 – Ethnic Factors in the Acceptability of Foreign Clinical Data

---

Questions and Answers

Q11: ***There seems to be an impression that the E5 bridging study would always be conducted after data in the original region is complete. Is this correct?***

It may be desirable in certain situations to achieve the goal of bridging by conducting a multi-regional trial under a common protocol that includes sufficient numbers of patients from each of multiple regions to reach a conclusion about the effect of the drug in all regions. Please provide points to consider in designing, analyzing and evaluating such a multi-regional trial.

A11: Bridging data should allow for extrapolation of data from one region to another. Although E5 speaks generally to extrapolation of data to a new region, E5 was not intended to suggest that the bridging study should necessarily follow development in another region. In the answer to Q1, it is made clear that it is also possible to include earlier studies conducted in several regions in a global drug development program so that bridging data might become available sooner. This can expedite completion of a global clinical development program and facilitate registration in all regions. A bridging study therefore can be done at the beginning, during or at the end of a global development program. For a multi-regional trial to serve as a bridging study for a particular region, it would need to have persuasive results in that region, because it is these regional results that can convince the regulators in that region that the drug is effective, and can "bridge" the results of trials in other regions in the registration application.

Guidance for Industry

E5 – Ethnic Factors in the Acceptability of Foreign Clinical Data

---

Questions and Answers

A multi-regional trial for the purpose of bridging could be conducted in the context of a global development program designed for near simultaneous world-wide registration. The objectives of such a study would be: (1) to show that the drug is effective in the region and (2) to compare the results of the study between the regions with the intent of establishing that the drug is not sensitive to ethnic factors. The primary endpoint(s) of the study should be defined and acceptable to the individual regions and data on all primary endpoints should be collected in all regions under a common protocol. In instances where the primary endpoints to be used by the regions are different, data for comparison purposes on all primary endpoints should be collected in all regions.

Guidance for Industry

E5 – Ethnic Factors in the Acceptability of Foreign Clinical Data

Questions and Answers

For a study intended to serve as a bridging study, the following points should be considered:

Planning

The multi-regional trial would have to satisfy requirements of the region where the application is to be filed with respect to design and analysis (see answer to Q1). In general, a multi-regional study should be designed with sufficient numbers of subjects so that there is adequate power to have a reasonable likelihood of showing an effect in each region of interest. Minor differences in design (e.g., age inclusion criteria, concomitant medication, etc.) may be acceptable and prior discussion with regulatory agencies is encouraged. For safety evaluation, it is important to make as uniform as possible the method for collection and assessment of safety information among regions.

Analysis

Given the goal of the multi-regional bridging study, it is critical to provide efficacy and safety results by region, with attention given to the usual analyses (e.g., demographic and baseline variables, patient disposition). It will be of interest also to examine consistency of effects across regions. In a dose response study, it will be especially important to analyze dose response relationships for efficacy and safety both within the regions and across the regions.

Evaluation

It is difficult to generalize about what study results would be judged persuasive, as this is clearly a regional determination, but a "hierarchy of persuasiveness" can be described.

1. Stand Alone Regional Result

The most persuasive would be demonstration of the effect in the entire study, with the results of each region of interest also demonstrating a statistically significant result. It will also be important to compare results across regions.

2. No Significant Regional Result But Similar Results Across Regions

With an effect demonstrated in the entire study, an analysis of results by region might not show a significant result in a region of interest but the data might nonetheless be persuasive to regulators in that region. Consistent trends in endpoint(s) intended for comparison across the regions or, in the case of a dose-response study, similar dose-response relationships across regions, might support an argument that the drug is not sensitive to intrinsic or extrinsic ethnic factors. Other data, for example, from approved drugs in the same class within region(s) could support such a bridging conclusion.

Some Examples

Cloprigrel (Plavix)

- Genetic interaction ?
  - Is it real – type of evidence needed – association vs. causality
- Drug-drug interaction
  - Neutralizing the important clinical response of one drug with use of another drug
- Regional treatment effect sizes and differences ?

Inference from a single exposure cohort

Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy of Clopidogrel Therapy

Alan R. Shuldiner, MD
Jeffrey R. O'Connell, DPhil
Kevin P. Bliden, BS
Amish Gandhi, MD
Kathleen Ryan, MPH
Richard B. Horenstein, MD
Coleen M. Damcott, PhD
Ruth Pakyz, BS
Udaya S. Tantry, PhD
Quince Gibson, MBA

Context Clopidogrel therapy improves cardiovascular outcomes in patients with acute coronary syndromes and following percutaneous coronary intervention by inhibiting adenosine diphosphate (ADP)–dependent platelet activation. However, nonresponsiveness is widely recognized and is related to recurrent ischemic events.

Objective To identify gene variants that influence clopidogrel response.

Design, Setting, and Participants In the Pharmacogenomics of Antiplatelet Intervention (PAPI) Study (2006-2008), we administered clopidogrel for 7 days to 429 healthy Amish persons and measured response by ex vivo platelet aggregometry. A genome-wide association study was performed followed by genotyping the loss-of-function cytochrome P450 (CYP) *2C19*2* variant (rs4244285). Findings in the PAPI Study were extended by examining the relation of *CYP2C19*2* genotype to platelet function and cardiovascular outcomes in an independent sample of 227 patients undergoing percutaneous coronary intervention.

Conclusion *CYP2C19*2* genotype was associated with diminished platelet response to clopidogrel treatment and poorer cardiovascular outcomes.

JAMA. 2009;302(8):849-858 www.jama.com

Inference from a single cohort exposure

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Cytochrome P-450 Polymorphisms and Response to Clopidogrel

Jessica L. Mega, M.D., M.P.H., Sandra L. Close, Ph.D., Stephen D. Wiviott, M.D., Lei Shen, Ph.D., Richard D. Hockett, M.D., John T. Brandt, M.D., Joseph R. Walker, Pharm.D., Elliott M. Antman, M.D., William Macias, M.D., Ph.D., Eugene Braunwald, M.D., and Marc S. Sabatine, M.D., M.P.H.

CONCLUSIONS

Among persons treated with clopidogrel, carriers of a reduced-function *CYP2C19* allele had significantly lower levels of the active metabolite of clopidogrel, diminished platelet inhibition, and a higher rate of major adverse cardiovascular events, including stent thrombosis, than did noncarriers.

A different conclusion from the randomized comparison (sub-samples from two RCT's)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

This article (10.1056/NEJMoa1008410) was published on August 29, 2010, at NEJM .org.

Effects of CYP2C19 Genotype on Outcomes of Clopidogrel Treatment

Guillaume Paré, M.D., Shamir R. Mehta M.D., Salim Yusuf, D.Phil., F.R.C.P.C., Sonia S. Anand, M.D., Ph.D., Stuart J. Connolly, M.D., Jack Hirsh, M.D., Katy Simonsen, Ph.D., Deepak L. Bhatt, M.D., M.P.H., Keith A.A. Fox, M.D., and John W. Eikelboom, M.D.

CONCLUSIONS

Among patients with acute coronary syndromes or atrial fibrillation, the effect of clopidogrel as compared with placebo is consistent, irrespective of *CYP2C19* loss-of-function carrier status. (Funded by Sanofi-Aventis and Bristol-Myers Squibb; ClinicalTrials .gov number, NCT00249873.)

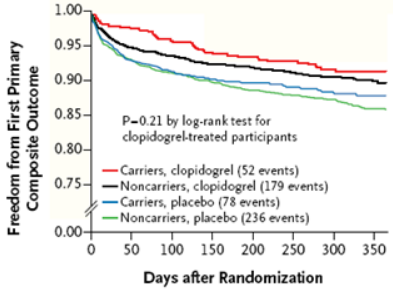
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effects of CYP2C19 Genotype on Outcomes of Clopidogrel Treatment

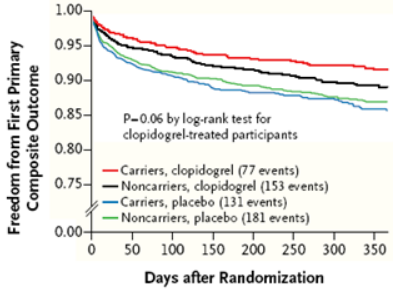
Guillaume Paré, M.D., Shamir R. Mehta M.D., Salim Yusuf, D.Phil., F.R.C.P.C., Sonia S. Anand, M.D., Ph.D., Stuart J. Connolly, M.D., Jack Hirsh, M.D., Katy Simonsen, Ph.D., Deepak L. Bhatt, M.D., M.P.H., Keith A.A. Fox, M.D., and John W. Eikelboom, M.D.

A First Primary Composite Outcome According to Loss-of-Function Allele Carrier Status



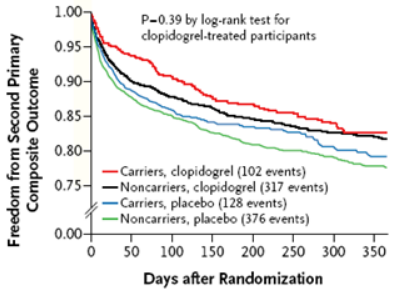
No. at Risk								
Carriers, clopidogrel	651	632	608	545	484	425	358	297
Noncarriers, clopidogrel	1886	1778	1723	1541	1352	1191	960	804
Carriers, placebo	674	626	609	551	483	423	356	281
Noncarriers, placebo	1819	1686	1634	1456	1259	1103	922	774

B First Primary Composite Outcome According to Gain-of-Function Allele Carrier Status



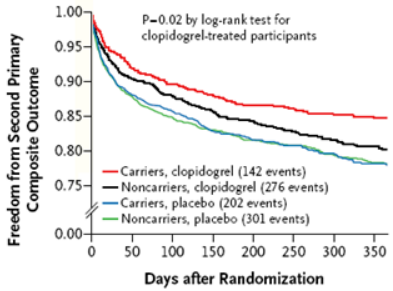
No. at Risk								
Carriers, clopidogrel	1001	960	932	828	715	640	513	428
Noncarriers, clopidogrel	1536	1451	1400	1258	1122	977	805	673
Carriers, placebo	1004	926	899	789	678	596	494	407
Noncarriers, placebo	1489	1386	1343	1218	1066	931	783	647

C Second Primary Composite Outcome According to Loss-of-Function Allele Carrier Status



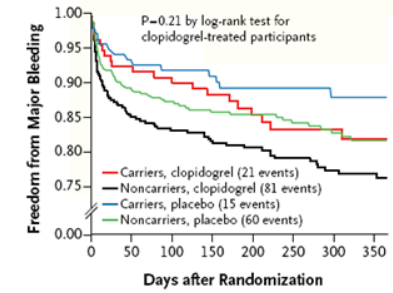
No. at Risk								
Carriers, clopidogrel	651	609	578	510	450	397	332	270
Noncarriers, clopidogrel	1886	1695	1622	1444	1248	1091	871	732
Carriers, placebo	674	599	574	514	446	389	323	248
Noncarriers, placebo	1819	1595	1530	1349	1152	1004	832	699

D Second Primary Composite Outcome According to Gain-of-Function Allele Carrier Status



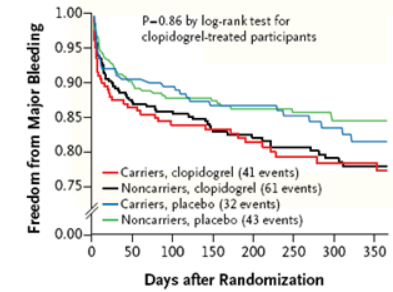
No. at Risk								
Carriers, clopidogrel	1001	919	882	780	664	592	467	390
Noncarriers, clopidogrel	1536	1386	1319	1174	1035	897	736	612
Carriers, placebo	1004	883	851	740	629	551	453	374
Noncarriers, placebo	1489	1310	1251	1122	970	842	701	572

E Freedom from Major Bleeding According to Loss-of-Function Allele Carrier Status



No. at Risk								
Carriers, clopidogrel	651	630	613	548	485	426	362	297
Noncarriers, clopidogrel	1886	1782	1726	1549	1365	1198	971	817
Carriers, placebo	674	647	633	577	504	444	371	294
Noncarriers, placebo	1819	1737	1693	1513	1314	1146	954	803

F Freedom from Major Bleeding According to Gain-of-Function Allele Carrier Status



No. at Risk								
Carriers, clopidogrel	1001	957	929	828	718	637	512	427
Noncarriers, clopidogrel	1536	1456	1411	1269	1133	988	821	687
Carriers, placebo	1004	957	938	827	715	625	517	429
Noncarriers, placebo	1489	1427	1388	1264	1106	967	807	667

Drug – Drug Interactions

that impact the efficacy or safety of a treatment

Information for Healthcare Professionals: Update to the labeling of Clopidogrel Bisulfate (marketed as Plavix) to alert healthcare professionals about a drug interaction with omeprazole (marketed as Prilosec and Prilosec OTC) [11/17/2009]

FDA is alerting the public to new safety information concerning an interaction between clopidogrel (Plavix), an anti-clotting medication, and omeprazole (Prilosec/Prilosec OTC), a proton pump inhibitor (PPI) used to reduce stomach acid. New data show that when *clopidogrel and omeprazole are taken together, the effectiveness of clopidogrel is reduced*. Patients at risk for heart attacks or strokes who use clopidogrel to prevent blood clots will not get the full effect of this medicine if they are also taking omeprazole. The updated label for clopidogrel will contain details of new studies submitted by Sanofi-Aventis and Bristol-Myers Squibb, the manufacturer of Plavix (clopidogrel).

Example of differential treatment effects - what to make of it -In a multi-regional study

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 SEPTEMBER 10, 2009 VOL. 361 NO. 11

Ticagrelor versus Clopidogrel in Patients with Acute Coronary Syndromes

Questions

Ticagrelor

July 28, 2010

Department Of Health And Human Services
Public Health Service
Food and Drug Administration

The Advisory Committee is asked to opine on the approvability of ticagrelor to reduce thrombotic events in patients with acute coronary syndromes or myocardial infarction, whether treatment is intended to be medical management or percutaneous coronary intervention (PCI).

The support for this claim comes primarily from PLATO, a randomized, event-driven double-blind comparison of ticagrelor (180 mg loading dose plus 90 mg twice daily) and clopidogrel (300 or 600 mg loading dose plus 75 mg daily), on a background of aspirin (anywhere from 160 to 500 mg loading plus 75 to 325 mg daily). The primary end point was time to first event of cardiovascular mortality, myocardial infarction, or stroke, tested with α=0.05 (adjusted for one interim analysis). Overall results were as follows:

	Clopidogrel n=9291	**Ticagrelor n=9333**	**HR**
CV death, MI, stroke	10.9%	9.3%	0.84 (0.77-0.92)
MI	6.4%	5.4%	0.84 (0.75-0.95)
CV death	4.8%	3.8%	0.79 (0.69-0.91)
Stroke	1.1%	1.3%	1.17 (0.91-1.52)

4. Do you believe the difference in clinical outcomes between the US and the rest of the world was attributable …

   4.1. … the play of chance? There is only one country out of 43 whose results fall outside the 95% confidence limits for a region having the observed number of events. If you think that chance is the most likely explanation, are you sufficiently sure of that to take the overall results to be applicable to the US?

   4.2. … a difference in dosing of aspirin, which was generally higher in the US? If so…

      4.2.1. Aspirin dose was one factor among dozens explored. How do you adjust for such multiplicity?

      4.2.2. How compelling are the external data that the dose of aspirin makes any difference in prevention of thrombotic events?

      4.2.3. How do you explain the apparently different effect of aspirin dose on ticagrelor and clopidogrel?

      4.2.4. If you think that aspirin dose is the most likely explanation for the discouraging results in the U.S., do you feel sufficiently sure that when administered with a low dose of aspirin, Brilinta will provide a clinical advantage over clopidogrel in the U.S. population?

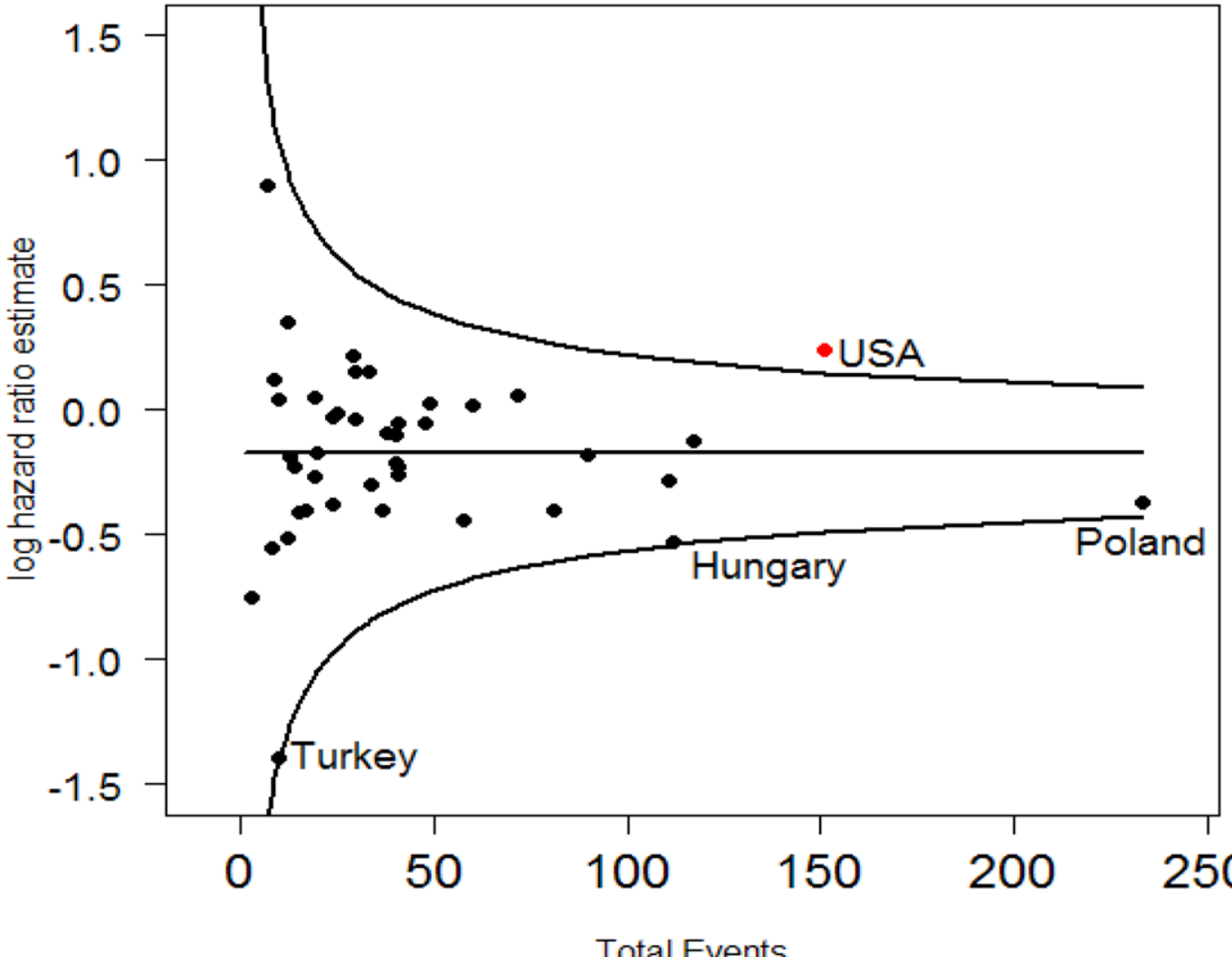
   4.3. … some other identifiable factor?

   4.4. … some unidentified set of population and care factors?

	Ticagrelor (n/N)	Clopidogrel (n/N)	HR (95% CI)
PLATO Overall N=18,624	9.8% (864/9333)	11.7% (1014/9291)	**0.84** (0.77, 0.93)
Non-US n=17,211	9.6% (780/8626)	11.8% (947/8585)	**0.81** (0.74, **0.90**)
US n=1,413	12.6% (84/707)	10.1% (67/706)	**1.27** (**0.92**, 1.75)

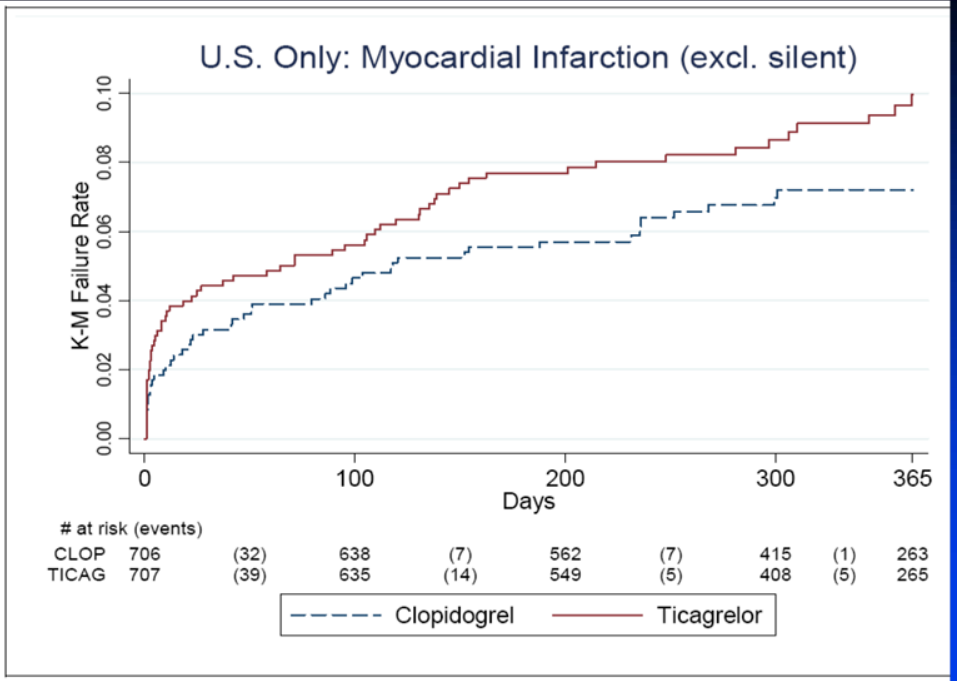
- 95% CIs of the US and non-US subgroups do not overlap
- In the US, clopidogrel did 'better' and ticagrelor did 'worse'

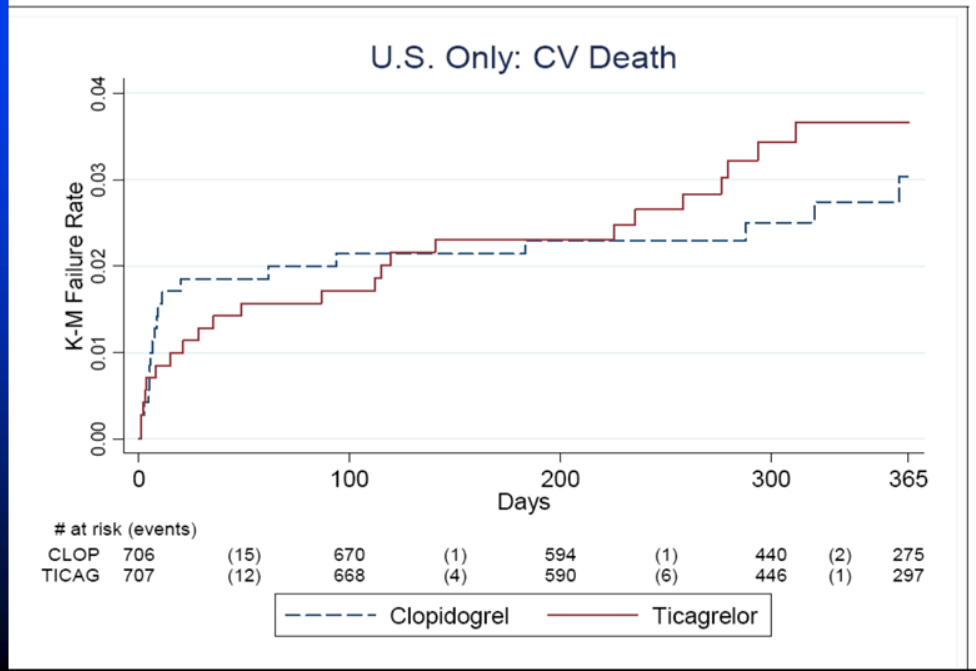
Funnel Plot: US is an outlier

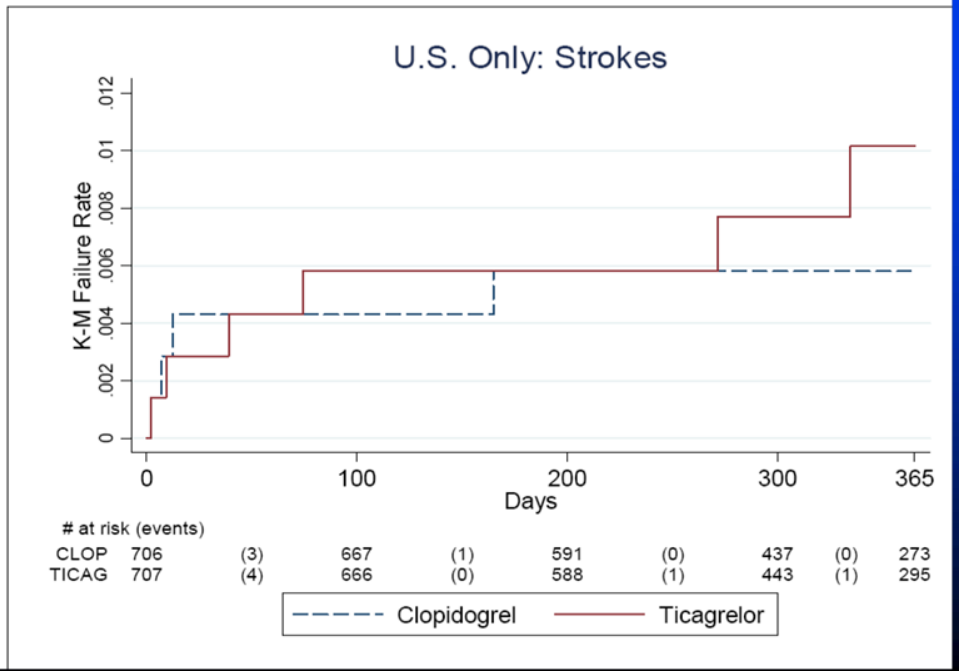


US ONLY

MI: 1.38 (0.95, 2.01)
CV Death: 1.26 (0.69, 2.31)
Strokes: 1.75 (0.51, 5.97)







Possible Explanations for US versus Non-US Difference

- Play of chance
- Concurrent ASA
- Other factors

Study Design Issues

- Enrichment designs to select or refine potential responder population
- Fixed vs. adaptive designs that adapt later stage populations to earlier stage findings
- Exploration vs. confirmation

Exploratory studies – I-SPY 2

I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy

AD Barker[1], CC Sigman[2], GJ Kelloff[1], NM Hylton[3], DA Berry[4] and LJ Esserman[3]

Received 10 February 2009; accepted 30 March 2009; advance online publication 13 May 2009. doi:10.1038/clpt.2009.68

CLINICAL PHARMACOLOGY & THERAPEUTICS

I-SPY 2 (investigation of serial studies to predict your therapeutic response with imaging and molecular analysis 2) is a process targeting the rapid, focused clinical development of paired oncologic therapies and biomarkers. The framework is an adaptive phase II clinical trial design in the neoadjuvant setting for women with locally advanced breast cancer. I-SPY 2 is a collaborative effort among academic investigators, the National Cancer Institute, the US Food and Drug Administration, and the pharmaceutical and biotechnology industries under the auspices of the Foundation for the National Institutes of Health Biomarkers Consortium.

Confirmatory studies

- Fixed sample size
- Adaptive designs

The range of applications

- Co-development of a drug and a device (diagnostic) to identify the patients to benefit
- Need for characterizing the sensitivity , specificity; positive and negative predictive value of the classifier – when and how
- Inference on an individual patient outcome vs. inference on the group or risk set one belongs to
- Labeling a product for use by a particular subpopulation

The classifier (personalization) metric

- What do you need to know about it
- When do you need to know about it
- If available it may be so, for different reasons (indications)
  - Some examples

Some examples of the use of diagnostic tests

- Her 2 IHC, HER2 Fish/CISH – appear in the product label because they were used to select patients in the clinical trials
  - The indication for the HER2 test does not say that it is a predictive marker (ie. Predicts differential treatment response)
- MammaPrint, an approved test as a 'prognostic' marker for breast cancer patients for risk of distant metastases
- Some tests are approved for screening: PSA, CEA
- Most oncology products approved for a broader population – few target therapies developed yet

K-ras – What did we learn

- Evaluating the consistency of effects across multiple studies and within drug class
- Impact of convenience samples for the Kras classification
- Retrospective vs. prospective analysis strategies
- How to take advantage of pre-treatment/randomization, baseline screening

Cetuximab Trials –Class safety labeling revision : lack of benefit in the K-ras mutant mCRC population

Clinical Trial	Line	Add'al Therapy	1° Endpt	Met 1° Endpoint	ITT Patients Tested for KRAS			Assay
					n	ITT	% of ITT	
CRYSTAL EMR 62202-013	1st	FOLFIRI	PFS	YES p = 0.048	540	1198	45	PCR based
NCIC-017 CA225025	3rd	BSC	OS	YES p = 0.005	394	572	69	sequencing
EPIC CA225006	2nd	irinotecan	OS	NO p = 0.71	300	1298	23	sequencing
OPUS EMR 62 202-047	1st	FOLFOX	RR	NO p = 0.06	233	337	69	PCR based

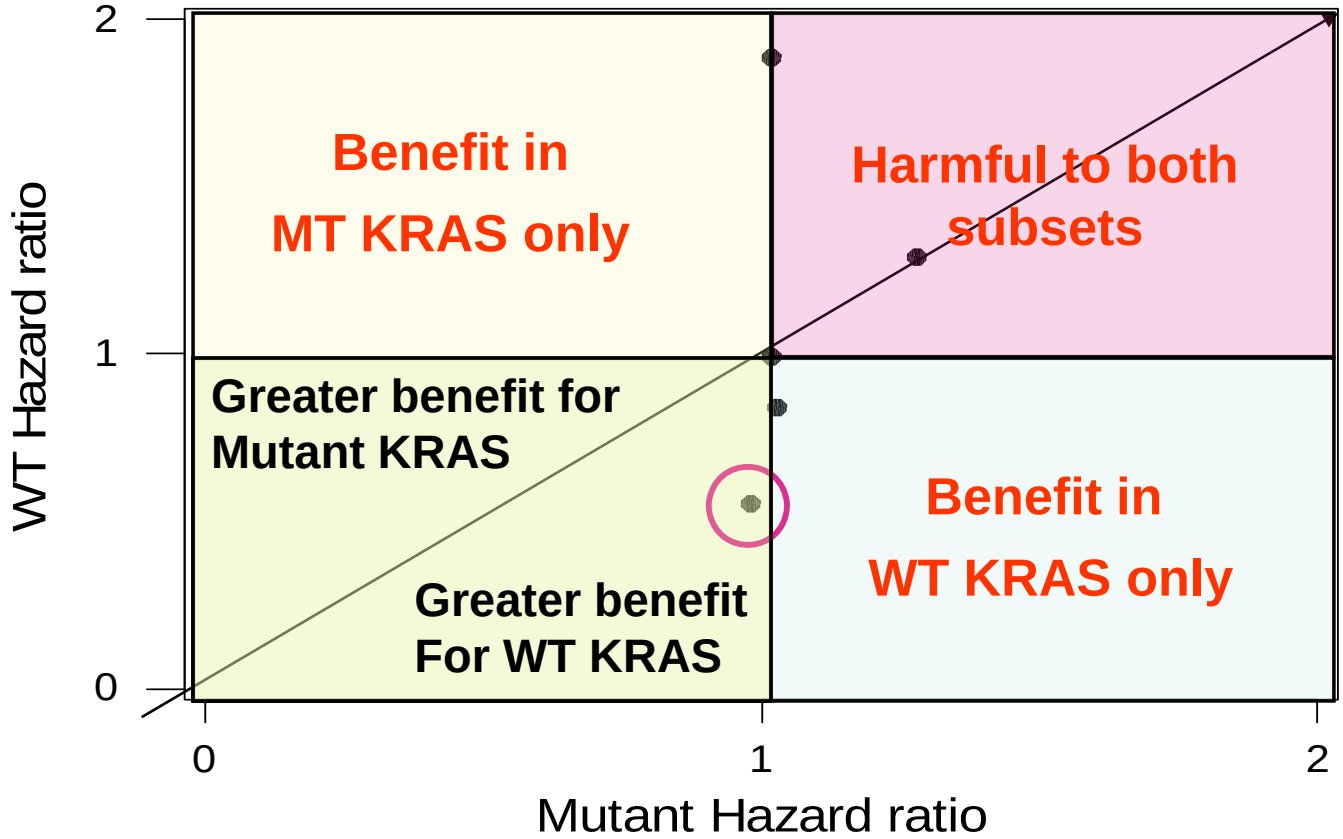
Panitumumab Trials - Class safety labeling revision : lack of benefit in the K-ras mutant mCRC population

Clinical Trial	Line	Add'nl Therapy	1° Endpt	Met 1° Endpoint	ITT Patients Tested for KRAS			Assay
					n	ITT	% of ITT	
20020408	3rd	BSC	PFS	YES p < 0.0001	427	463	92	PCR based
PACCE 20040249	1st	chemo/ bev	PFS	NO Inferior P = 0.002	863	1053	82	PCR based

Overall survival

- The following graph provides a summary of overall survival for the five studies having overall survival comparisons for the WT KRAS and mutant KRAS subgroups
- Hazard ratios are used for overall survival.
- Points above the line correspond to larger effects for Cetuximab or Panitumumab for the mutant KRAS “subgroup” than for the wild-type “subgroup”
- Points below the line correspond to larger effects for Cetuximab or Panitumumab for the wild-type “subgroup” than for the mutant KRAS “subgroup”

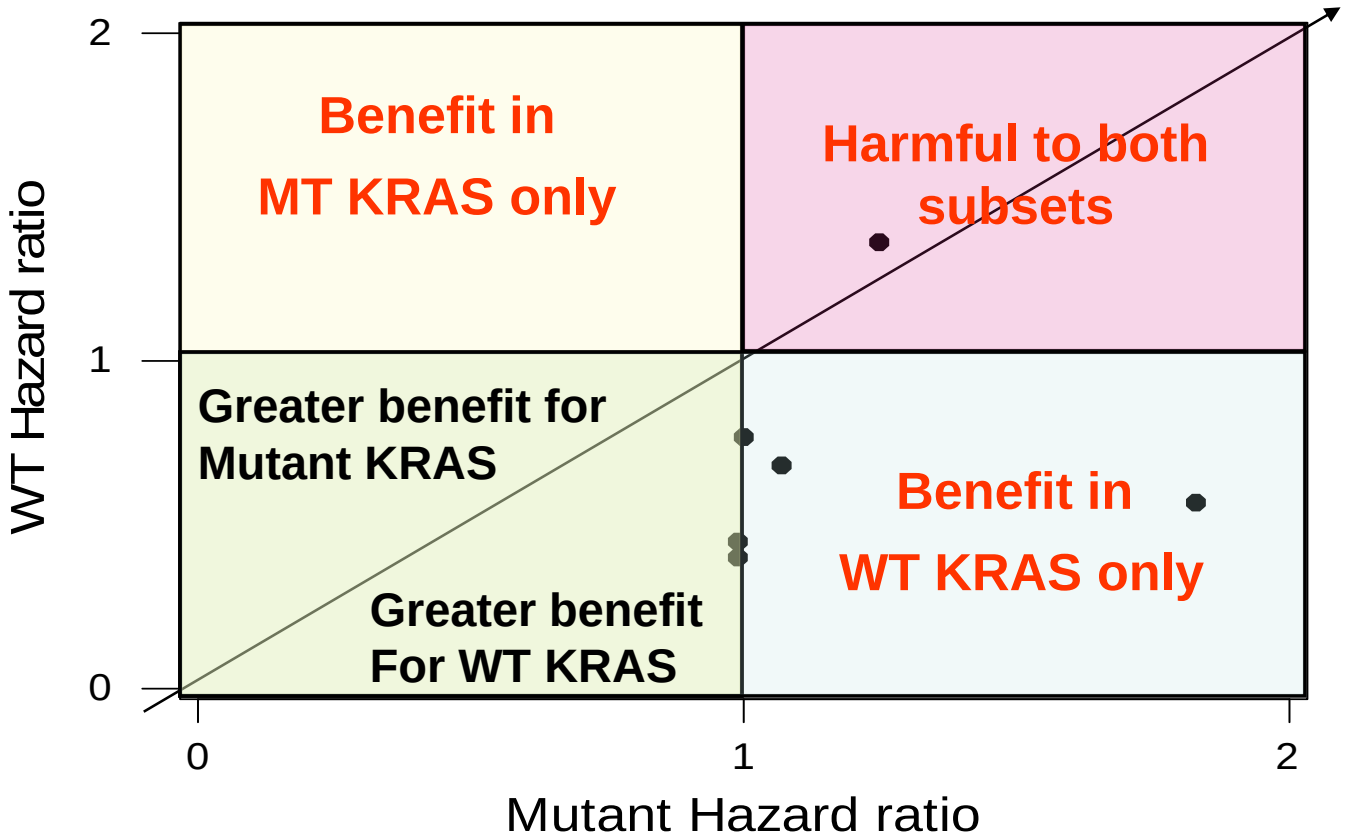
Overall Survival: three trials showed no benefit or harmful effect to both subsets; only the circled trial shows clear benefit in WT KRAS patients only



Progression-free survival

- The following graph provides a summary for the six studies of the progression-free survival comparisons for the WT KRAS and mutant KRAS subgroups
- Hazard ratios are used for progression-free survival.
- Points above the line correspond to larger effects for Cetuximab or Panitumumab for the mutant KRAS “subgroup” than for the wild-type “subgroup”
- Points below the line correspond to larger effects for Cetuximab or Panitumumab for the wild-type “subgroup” than for the mutant KRAS “subgroup”

PFS: five trials show benefit in WT KRAS only, one trial shows harmful effect to both subsets

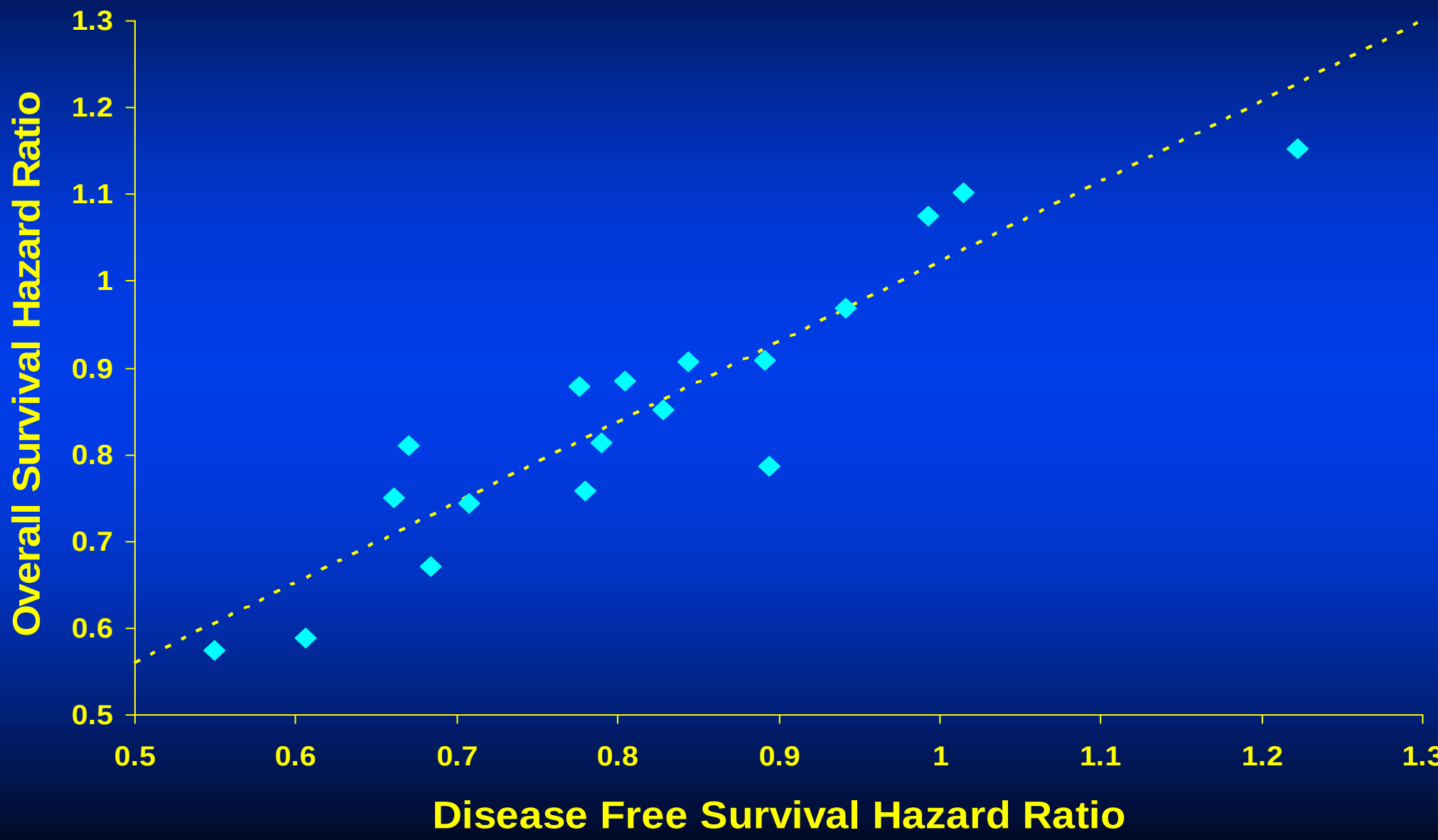


Moving from the retrospective assessment to prospective assessment – and from convenience samples to full ascertainment

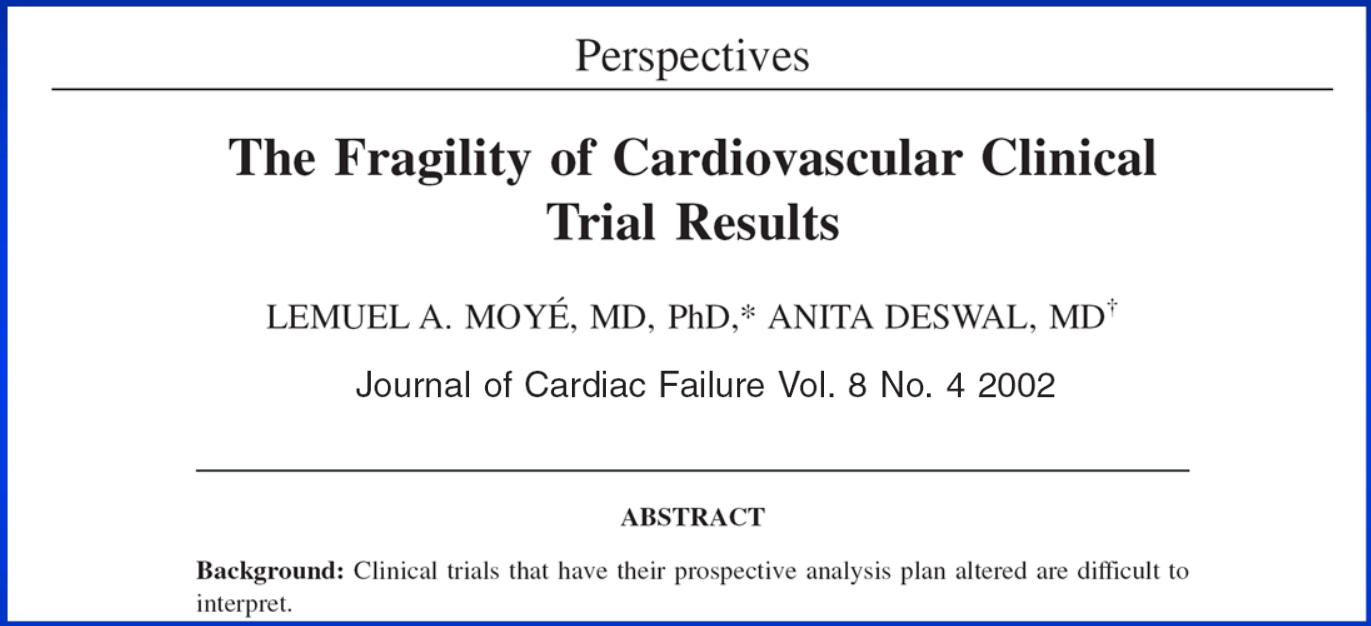
- The KRAS mutation experience (metastatic colorectal cancer)
- The BRAF mutation experience (malignant melanoma)

Some lessons learned from surrogate marker validation and unexpected findings in small subgroups

Colorectal Adj: Hazard Ratios for DFS vs. Overall Survival



Trials whose findings were reversed upon completion of a second study planned to specifically test the hypothesis generated in the first study



Perspectives

The Fragility of Cardiovascular Clinical Trial Results

LEMUEL A. MOYÉ, MD, PhD,* ANITA DESWAL, MD†

Journal of Cardiac Failure Vol. 8 No. 4 2002

ABSTRACT

Background: Clinical trials that have their prospective analysis plan altered are difficult to interpret.

Why ?

The analysis plan , in the initial study, changed after seeing the data: it placed new emphasis on a subgroup finding, or on a secondary endpoint raised in prominence leading to false discoveries that were not replicated

3 Examples: Vesnarinone, Amlodipine, Losartan

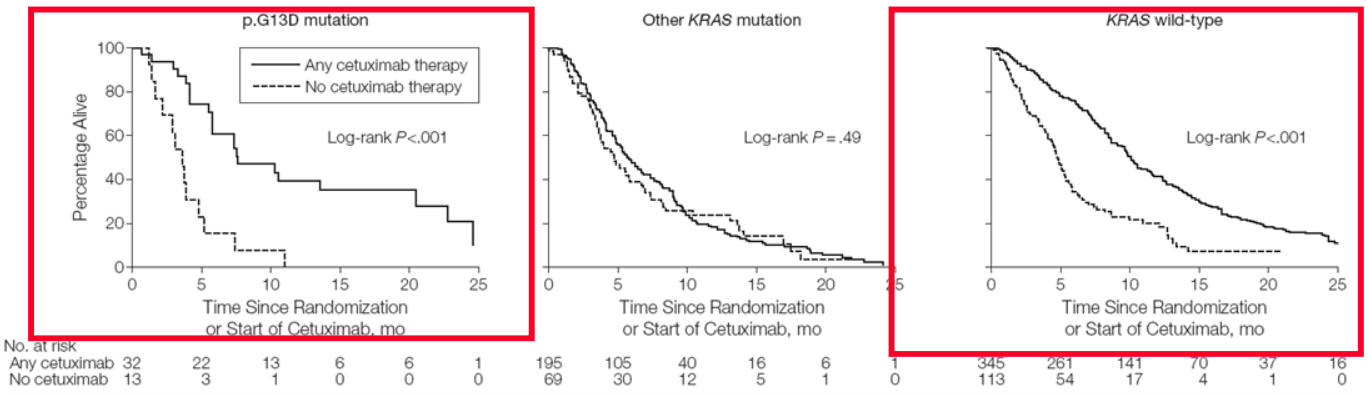
We should be concerned not only about Type 1 error control – true positives

- Replication of the treatment effect in the classified subset in a separate independent study
  - Especially when a marker has no biological / mechanistic interpretation
- Probability that the treatment effect in the marker subgroup is a *true positive – especially when the subset is relatively small, stratified randomization is not employed, and perhaps 100 % of the ITT population does not have classifier status ascertained*

How large should the 'off' group be – subsets who may benefit within mutant subset (the specificity of the classifier is an issue)

Association of *KRAS* p.G13D Mutation With Outcome in Patients With Chemotherapy-Refractory Metastatic Colorectal Cancer Treated With Cetuximab

Figure 1. Overall Survival: Predictive Analysis by *KRAS* Status for Patients Receiving Any Cetuximab-Based Therapy vs No Cetuximab



The no cetuximab group for all patients from the pooled data set is the best supportive care group from the CO.17 trial.

Conclusions In this analysis, use of cetuximab was associated with longer overall and progression-free survival among patients with chemotherapy-refractory colorectal cancer with p.G13D-mutated tumors than with other *KRAS*-mutated tumors. Evaluation of cetuximab therapy in these tumors in prospective randomized trials may be warranted.

JAMA. 2010;304(16):1812-1820 www.jama.com

How probable are prognostic factor imbalances ? It Depends

(Implications for minimum marker subgroup sample sizes to minimize bias)

- Full ITT population - factor ascertained on everyone in the RCT
  - Depends upon sample size in each treatment group within each factor (genomic + or -)
- Convenience samples - factor is ascertained on a non-randomized subset of subjects, in each treatment group – Imbalance in prevalence of prognostic markers in each non-randomized subgroup and imbalance in prevalence of marker status can introduce biases in the data

Table 1. Probability of observed imbalance between two treatment groups: a binary prognostic factor*

True Prevalence	N=20/arm	N=50/arm	N=100/arm
10%	0.0631	0.0017	0.0000
20%	0.1636	0.0173	0.0006
30%	0.2258	0.0377	0.0026
40%	0.2582	0.0519	0.0048
50%	0.2682	0.0569	0.0057

* imbalance is defined as a 20% observed difference, Cui et al. (2002).

Table 2. Probability of imbalance in prognostic factor for certain sample sizes and percent imbalance

Prevalence	N=1350 d = 5%	N=350 d = 10%	N=150 d = 15%
10%	0.0000	0.0000	0.0000
20%	0.0012	0.0011	0.0012
30%	0.0046	0.0044	0.0045
40%	0.0080	0.0077	0.0079
50%	0.0094	0.0091	0.0093

d: % observed imbalance between the treated group and the comparator group; Wang, O'Neill, Hung (2010).

A RCT to demonstrate minimizing risk through effective screening
The Abacavir 'PREDICT -1' trial

- Same treatment in both randomized groups
- Treatment groups differed by screening strategy and the entrance criteria into the trial
  - HLA-B*5701 screening - exclude positive subjects in one of the randomized arms
  - Goal: demonstrate screening reduces incidence of serious adverse event
  - **Provides estimates of Sensitivity and specificity for the classifier**

N Engl J Med 2008;358:568-79.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

HLA-B*5701 Screening for Hypersensitivity to Abacavir

Simon Mallal, M.B., B.S., Elizabeth Phillips, M.D., Giampiero Carosi, M.D., Jean-Michel Molina, M.D., Cassy Workman, M.B., B.S., Janez Tomažič, M.D., Eva Jägel-Guedes, M.D., Sorin Rugina, M.D., Oleg Kozyrev, M.D., Juan Flores Cid, M.D., Phillip Hay, M.B., B.S., David Nolan, M.B., B.S., Sara Hughes, M.Sc., Arlene Hughes, Ph.D., Susanna Ryan, Ph.D., Nicholas Fitch, Ph.D., Daren Thorborn, Ph.D., and Alastair Benbow, M.B., B.S., for the PREDICT-1 Study Team*

The NEW ENGLAND JOURNAL of MEDICINE

N Engl J Med 2008;358:568-79.

ORIGINAL ARTICLE

HLA-B*5701 Screening for Hypersensitivity to Abacavir

Simon Mallal, M.B., B.S., Elizabeth Phillips, M.D., Giampiero Carosi, M.D., Jean-Michel Molina, M.D., Cassy Workman, M.B., B.S., Janez Tomažič, M.D., Eva Jägel-Guedes, M.D., Sorin Rugina, M.D., Oleg Kozyrev, M.D., Juan Flores Cid, M.D., Phillip Hay, M.B., B.S., David Nolan, M.B., B.S., Sara Hughes, M.Sc., Arlene Hughes, Ph.D., Susanna Ryan, Ph.D., Nicholas Fitch, Ph.D., Daren Thorborn, Ph.D., and Alastair Benbow, M.B., B.S., for the PREDICT-1 Study Team*

METHODS

This double-blind, prospective, randomized study involved 1956 patients from 19 countries, who were infected with human immunodeficiency virus type 1 and who had not previously received abacavir. We randomly assigned patients to undergo prospective HLA-B*5701 screening, with exclusion of HLA-B*5701–positive patients from abacavir treatment (prospective-screening group), or to undergo a standard-of-care approach of abacavir use without prospective HLA-B*5701 screening (control group). All patients who started abacavir were observed for 6 weeks. To immunologically confirm, and enhance the specificity of, the clinical diagnosis of hypersensitivity reaction to abacavir, we performed epicutaneous patch testing with the use of abacavir.

The study confirmed the hypothesis that screening will reduce a severe adverse reaction

Table 2. Incidence of Hypersensitivity Reaction to Abacavir.*

Hypersensitivity Reaction	Prospective Screening	Control	Odds Ratio (95% CI)*	P Value
	no. of patients/total no. (%)			
Clinically diagnosed				
Total population that could be evaluated	27/803 (3.4)	66/847 (7.8)	0.40 (0.25–0.62)	P<0.001
White subgroup	24/679 (3.5)	61/718 (8.5)	0.38 (0.23–0.62)	P<0.001
Immunologically confirmed				
Total population that could be evaluated	0/802	23/842 (2.7)	0.03 (0.00–0.18)	P<0.001
White subgroup	0/679	22/713 (3.1)	0.03 (0.00–0.19)	P<0.001

* P values, odds ratios, and 95% confidence intervals (CIs) were calculated by means of logistic-regression analysis and adjusted for self-reported race (white vs. nonwhite), history of receipt of antiretroviral therapy (none vs. any), introduction of a new nonnucleoside reverse-transcriptase inhibitor (yes or no), and concurrent use or nonuse of a protease inhibitor. The model-based incidences of clinically diagnosed hypersensitivity reaction to abacavir in the total population that could be evaluated and in the white subgroup were 3.3% and 3.5%, respectively, for the prospective-screening group and 7.9% and 8.6%, respectively, for the control group. The white subgroup included the two and three patients reporting both categories of white ancestry in the prospective-screening group and the control group, respectively. The odds ratios for immunologically confirmed hypersensitivity reaction to abacavir were obtained by means of exact methods, owing to the absence of immunologically confirmed hypersensitivity reaction in the prospective-screening group. The model (involving a median, unbiased estimate of the odds ratio) estimated the odds of hypersensitivity reaction in the prospective-screening group versus the control group to be 1:33 ($1 \div 0.03 = 33$). (Although a simple point estimate of the odds ratio from the raw data yields a more intuitive value of 0, it also implies an infinite reduction in the odds, which is problematic for linear regression modeling in that it introduces error from division by 0.)

The study design also provided estimates of performance of the classifier – sensitivity and specificity

Table 4. Performance Characteristics of HLA-B*5701 Screening for Hypersensitivity Reaction to Abacavir in the Control Group.*

Subgroup	Positive for HLA-B*5701	Negative for HLA-B*5701	Total	Performance Characteristic for Hypersensitivity Reaction
		number of patients		*percent (95% CI)*
Clinically diagnosed hypersensitivity reaction				
Total population that could be evaluated				
Hypersensitivity reaction	30	36	66	Sensitivity: 45.5 (33.1–58.2)
No hypersensitivity reaction	19	762	781	Specificity: 97.6 (96.2–98.5) PPV: 61.2 (46.2–74.8) NPV: 95.5 (93.8–96.8)
White subgroup				
Hypersensitivity reaction	29	32	61	Sensitivity: 47.5 (34.6–60.7)
No hypersensitivity reaction	19	638	657	Specificity: 97.1 (95.5–98.3) PPV: 60.4 (45.3–74.2) NPV: 95.2 (93.3–96.7)
Immunologically confirmed hypersensitivity reaction				
Total population that could be evaluated				
Hypersensitivity reaction	23	0	23	Sensitivity: 100 (85.2–100)
No hypersensitivity reaction	25	794	819	Specificity: 96.9 (95.5–98.0) PPV: 47.9 (33.3–62.8) NPV: 100 (99.5–100)
White subgroup				
Hypersensitivity reaction	22	0	22	Sensitivity: 100 (84.6–100)
No hypersensitivity reaction	25	666	691	Specificity: 96.4 (94.7–97.6) PPV: 46.8 (32.1–61.9) NPV: 100 (99.4–100)

* The white subgroup included the two and three patients reporting both categories of white ancestry in the prospective-screening group and the control group, respectively. NPV denotes negative predictive value, and PPV positive predictive value.

Where might RCT's being going in the future
Prospective study design options

- A two stage design that reserves some type 1 error for testing a subgroup yet to be specified - (biological plausibility)
  - Fixed study design with no adaptation to increase samples size overall or in subgroups
- An adaptive study design that can increase sample size and pre-specifies the 'win criteria' or study 'success' criteria
- Also tests the efficacy of a classifier at the same time the prognostic effect is demonstrated

What do we need to know for a marker to be predictive of treatment effect (relative change in response)

- An unbiased comparison between the test treatment and control in each of the marker subgroups
  - Unbiased generally requires a randomized subset of subjects in each of the marker categories, not a convenience sample of subjects with marker status available

Performance of assays for marker classification

- What are the minimum performance characteristics (e.g., sensitivity, specificity, reproducibility) of the assay used to classify patient subgroups and what are the consequences of that performance for making correct inferences from the study
  - KRAS vs EGFR vs breast cancer assay
- In general, 'classifier' performance and marker prevalence (mix) may explain study to study heterogeneity and differences in results

Other designs and considerations

PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/pst.300

WILEY InterScience® DISCOVER SOMETHING GREAT

Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset[‡,§]

MAIN PAPER

Sue-Jane Wang[1,*,†], Robert T. O'Neill[1] and H. M. James Hung[2]
[1] *Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA*
[2] *Division of Biometrics I/OB, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA*

CLINICAL TRIALS *PRESENTATION*

Clinical Trials 2010; **7**: 525–536

Statistical considerations in evaluating pharmacogenomics-based clinical effect for confirmatory trials

Sue-Jane Wang[a], Robert T O'Neill[a] and HM James Hung[b]

Perspective

Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials

Clinical Cancer Research *Vol. 10, 6759–6763, October 15, 2004*

Richard Simon and Aboubakar Maitournam

Biometric Research Branch, National Cancer Institute, Bethesda, Maryland

these targeted designs. As discussed in this article, v
the efficiency of targeted designs in comparison with
randomized designs with broader eligibility criteria.
ated efficiency in the context of a binary outcome

STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:329–339
Published online 18 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.1975

On the efficiency of targeted clinical trials

A. Maitournam and R. Simon*,†

Cancer Therapy: Clinical

Clinical Cancer Research

The Cross-Validated Adaptive Signature Design

Boris Freidlin[1], Wenyu Jiang[2], and Richard Simon[1]

Clin Cancer Res; 16(2); 691–8. ©*2010 AACR.*

Establish consensus on a good analysis plan for a retrospective evaluation

- Role of randomization to assure unbiased and fair comparisons
- Role of marker status classification - impact of convenience samples on biased estimates
- Marker classification performance
- Statistical control of false positive conclusions - how many hypotheses, which were primary, which failed
  - Accounting for multiplicity – on which outcomes (OS,PFS,RR)
- Data to generate the hypothesis vs. data to confirm the hypothesis
- Replication of evidence

Establish a framework for the level of rigor required

- Proof of concept - exploration of the association of a marker(s) with an outcome
  - In a cohort that is non exposed to test treatment - goal is prognostic factor
  - In a cohort exposed to test treatment - goal is a predictive factor
    - Could be both
- Proof of marker predictive treatment effects
  - Confirmatory clinical studies
    - Control of Type 1 error and minimizing bias
    - Replication - two or more studies showing the same consistent finding
    - PGx ascertainment on all randomized subjects with sufficient sample size in the minimum marker group to assure comparability of subject prognostic factors - addresses the confounding problem

Way forward

- Encourage RCT designs that evaluate subgroups in a more rigorous manner
- Studies may not necessarily be smaller if all marker subgroups are evaluated to identify best responders
- Guidances under development
  - Enrichment
  - Adaptive designs
  - Co-development